9Б

УДК 378.147:004.81

О.В. Бісікало
Вінницький державний аграрний університет, м. Вінниця, Україна
bisikalo@vsau.org

Методика побудови тезауруса навчальної системи на основі моделі образного мислення

Стаття присвячена створенню «розумного» контенту для електронних підручників. На основі вимог моделі образного мислення розглядаються принципи побудови тезауруса образів. Пропонується формальна реляційна модель тезауруса, що наповнюється в результаті обробки текстів з навчальним матеріалом, а також методика побудови відповідних програмно-педагогічних систем. Діючий прототип тезауруса образів програмно реалізовано на основі технології Python + SQLite.

Провідна роль комп'ютерних технологій в науково-технічній революції поки що не змінила існуючого положення речей, при якому значна частина людських знань зберігається в текстовій формі. Незалежно від типу носія інформації символьний-вербальний формат знань залишається стандартом для сучасної науки і освіти. Однією з проблем електронного дистанційного навчання є швидке послаблення уваги і, отже, знижений рівень сприйняття текстової інформації з екрана комп'ютера в порівнянні з традиційним підручником. Поліпшення якості моніторів поки що не повністю вирішує дану проблему і тим більше не враховує можливостей комп'ютерного представлення інформації.

Одним з основних наукових джерел комп'ютерної лінгвістики вважається психолінгвістика, фокус досліджень якої спрямовано на проблеми мовленнєвої діяльності людини і формування її мовної здатності. Визначення асоціативного значення слова ввів у сучасну наукову парадигму Дж. Діз, що став одним з найзначніших його дослідників у психолінгвістиці. У інтерпретації Діза асоціативне значення визначалося як найбільш адекватне психологічній структурі значення, як «потенційний розподіл відповідей на деяке слово-стимул». В процесі дослідження асоціацій Дж. Дізом детально вивчені імовірнісні характеристики асоціативних зв'язків слова з іншими словами за допомогою кластерного аналізу [1].

Багаторічні і широкомасштабні експерименти з дослідження особливостей текстотворчого потенціалу асоціативного поля слова А.А. Залевської та групи її співробітників розвинули підхід до аналізу значення слова, запропонований Дізом. На думку Залевської, цей підхід себе ще далеко не вичерпав, а перспективи його подальшого застосування визначаються можливістю інтеграції в більш загальну теорію специфіки індивідуального знання і принципів його функціонування [2].

Результати останніх досліджень в окресленому напрямку дозволяють говорити і про виникнення окремого напряму в психолінгвістиці – асоціативної лінгвістики (концепція Ю.М. Караулова), представленої в сучасному стані асоціативними лексикографією, лексикологією і граматикою. Згідно з цією концепцією, мова може бути представлена не тільки у вигляді системи відносин, не тільки як велика сукупність текстів, але і у вигляді асоціативно-вербальної мережі, що співвідноситься з мовленнєвою здатністю носія мови [3].

Мета даної статті полягає у покращенні сприйняття навчальної інформації в електронних підручниках на основі застосування поняття асоціативно-вербальної мережі (АВМ). Можливими шляхами досягнення мети принципово можуть бути:

– використання існуючих асоціативних словників, отриманих в результаті широкомасштабного вільного асоціативного експерименту;

– побудова власної АВМ на основі концепції моделювання образного мислення в онтогенезі, де окремі слова вважаються символьними позначеннями образів [4].

Перший підхід разом з певними труднощами у реалізації [5] призводить до отримання тих же самих узагальнених парадигматичних зв’язків дорослих носіїв мови. В межах другого підходу просте оповідне речення або синтагма тексту розглядається як аналог події, де образи об’єднуються між собою за допомогою асоціативних зв’язків [6]. Тоді введення синтагм навчального тексту у систему суттєво доповнює евристично створений гіпертекст з парадигматичними зв’язками – можливостями синтагматичного знання у вигляді додаткових асоціативних зв’язків між образами. Тому **постановкою завдання** вважатимемо формалізацію методики побудови тезауруса як словника образів предметної області виходячи з онтогенетичних вимог загальної моделі образного мислення.

Формалізація тезауруса

З точки зору організації роботи з системою першочерговою задачею, безумовно, є наповнення її інформацією у вигляді синтагм. Підхід, що пропонується, побудовано на моделюванні природного інтелекту шляхом самоспостереження і об’єднання відомих результатів таких наук, як фізіологія, психологія, лінгвістика, нейропсихологія, психолінгвістика тощо. Розглянемо побудову формального алгоритму внесення нової синтагми на основі принципів, що закладені в концептуальну модель образного мислення [4]. З ними не повинні входити в суперечність можливі формальні конструкції та технічні рішення системи.

В першу чергу необхідно врахувати принцип самонавчання живих носіїв інтелекту. Відомо, що немовля практично з нуля навчається всієї складності миру за простим принципом, сформульованим ще в стародавньому Римі як «поступовість і неперервність». Для сприйняття будь-якого нового знання потрібний хороший фундамент, тому словник образів необхідно «випестити» до необхідного рівня шляхом організаційної опіки від найбільш загальних образів та понять до спеціальних термінів. Інакше парадоксальною виглядає експертна система, що легко вирішує складні, але вузькоспеціалізовані завдання, проте безсило відступає перед «дитячими» питаннями.

Істотною відмінністю природної пам’яті від комп’ютерної є принципова відсутність операції стирання інформації. Якщо людина зустрічає новий образ, то вона ніколи його не проігнорує навіть тому, що в окремих випадках це може призвести до небезпечних наслідків. Точно також не можна видалити вже наявний образ, наприклад, негативний – випадки підсвідомого витіснення, як показав ще З. Фрейд, приводять до патології в психіці людини. Психоаналіз в якості методу лікування пропонує «витягування» образу в свідомість і «нарощування» на нього позитивної мережі зв’язків з іншими образами. Даний принцип організації пам’яті можна, декілька спрощуючи, сформулювати як «всіх впускати – нікого не випускати».

Третій з основних онтогенетичних постулатів моделі стосується механізму поповнення словника образів. Відомо, що при сприйнятті нового образу у людини на підсвідомому рівні включається орієнтовний рефлекс. Одним із завдань цього рефлексу є побудова мережі зв’язків нового образу з образами, що вже існують в пам’яті.

Іншими словами, відбуваються процеси розпізнавання або класифікації образу. Будь-який новий образ принципово не може бути незалежним, оскільки у момент введення в систему поєднується максимально можливою в даний момент часу кількістю зв'язків з іншими образами. Тим самим виявляється рекурентна сутність визначення образу – «скажи мені, хто твої друзі, і я скажу, хто ти».

Для реалізації формальної методики побудови тезауруса на основі запропонованих онтогенетичних принципів розглянемо алгебраїчну систему. До складу ***відношень***, які дозволяють представити систему у термінах реляційної моделі даних, будемо включати [6]:

$$RE = \left\{ \begin{matrix} \mathrm{Im}age, Assoc - Twice, Construct, Event, \\ Interrogative - \mathrm{Pr}onoun, Link, Text, Words, Role \end{matrix} \right\}, \quad (1)$$

де *Image* – ***образ***, що ставить у відповідність коду з множини *Bi-I* слова з підмножин, що складають множину *Image-Word*

$$\begin{matrix} \mathrm{Im}age \subset Bi - I \times Object - Quality \times Object \times \\ Notion \times Method \times Method - Quality \end{matrix}; \quad (2)$$

Assoc-Twice – ***асоціативна пара образів*** – пов'язує коди головного $Bi\text{-}I_1$ і підлеглого $Bi\text{-}I_2$ образів з кодом пари *Twice-Id* та значеннями сил прямого $Force^+$ та зворотного $Force^-$ зв'язку між ними

$$Assoc - Twice \subset Bi - I_1 \times Bi - I_2 \times Twice - Id \times Force^+ \times Force^- ; \quad (3)$$

за допомогою відношення *Assoc-Twice* можна поставити у відповідність кожному образу відсортований за силою зв'язку підсписок асоційованих з ним образів (хеш-таблиця), що і забезпечує ключову роль асоціативної пам'яті в образному мисленні;

Construct – ***складова синтагми*** – пов'язує код складової *Construct-Id* з кодом питального займенника *Pronoun-Id*, кодом пари *Twice-Id*, кодами головного $Word\text{-}Id_1$ та підлеглого $Word\text{-}Id_2$ слів пари, кодом події *Event-Id* та кодом ролі підлеглого слова в цій парі *Role-Id*

$$\begin{matrix} Construct \subset Construct - Id \times \mathrm{Pr}onoun - Id \times Twice - Id \times \\ Word - Id_1 \times Word - Id_2 \times Event - Id \times Role - Id \end{matrix}; \quad (4)$$

при створенні нового кортежу у відношенні *Construct* паралельно у відповідний кортеж відношення *Assoc-Twice* обов'язково додається 1 або до значення в атрибуті $Force^+$, або до значення в атрибуті $Force^-$ залежно від того, в якому порядку (прямому або у зворотному) було реалізовано даний асоціативний зв'язок;

Event – ***подія*** – пов'язує код події *Event-Id* з кодом синтагми *Bi-Sy*, кодом відповідного тексту *Text-Id* та власне текстом речення *Sentence*

$$Event \subset Event - Id \times Bi - Sy \times Text - Id \times Sentence ; \quad (5)$$

Interrogative-Pronoun – ***питальний займенник*** – пов'язує код займенника *Pronoun-Id* з кодом зв'язку *Link-Id*, до множини якого цей займенник належить, та словом *Pronoun*, що власне позначає займенник

$$Interrogative - \mathrm{Pr}onoun \subset \mathrm{Pr}onoun - Id \times Link - Id \times \mathrm{Pr}onoun ; \quad (6)$$

Link – ***асоціативний зв'язок*** – пов'язує код асоціативного зв'язку *Link-Id* з словом *Link-Type*, що позначає цей тип зв'язку та його вербальним описом *Specification*

$$Link \subset Link - Id \times Link - Type \times Specification ; \quad (7)$$

Text – ***текст*** – пов'язує унікальний код *Text-Id* з бінарним кодом тексту *Bi-Te*, його назвою *Title*, автором *Author* та часом *Time*

$$Text - RE \subset Text - Id \times Bi - Te \times Title \times Author \times Time ; \quad (8)$$

Words – ***вербальні ознаки образів*** – пов'язує унікальний код слова *Word-Id* з вербальним позначенням *Word* та кодом відповідного образа *Bi-I*

$$Words \subset Word - Id \times Word \times Bi - I ; \quad (9)$$

Role – ***роль образу у події*** – пов'язує унікальний код ролі *Role-Id* з словом *Role-Type*, що позначає роль образу у події

$$Role \subset Role - Id \times Role - Type\,. \tag{10}$$

Застосовані та описані вище назви атрибутів для відношень (2÷10) можна визначити наступним чином через основи алгебраїчної системи:

$$\mathrm{Re}-Word = \begin{Bmatrix} Sentence, \mathrm{Pr}\,onoun, Specification, \\ Title, Author, Word, Role - Type \end{Bmatrix}, \tag{11}$$

$$\begin{gathered} Twice - Id, Force^{+}, Force^{-}, Construct - Id \in Integer, \\ \mathrm{Pr}\,onoun - Id, Word - Id, Event - Id \in Integer, \\ Link - Id, Text - Id, Role - Id \in Integer. \end{gathered} \tag{12}$$

Як основи алгебраїчної системи для моделювання образного мислення використовуються слова *Word* та числа *Number*. В межах першої з основ *Word* варто розрізняти наступні види символьних послідовностей (слів):

$$Word = \{\mathrm{Im}\,age - Word, Link - Type, \mathrm{Re}-Word\}, \tag{13}$$

де слова, що позначають образи *Image-Word*, будемо класифікувати в залежності від їх ролі в синтагмі, а саме *Object-Quality* – якість об'єкта; *Object* – об'єкт; *Notion* – поняття; *Method* – метод; *Method-Quality* – якість методу:

$$\mathrm{Im}\,age - Word = \begin{Bmatrix} Object - Quality, Object, Notion, \\ Method, Method - Quality \end{Bmatrix}. \tag{14}$$

Словами *Link-Type* будемо позначати ролі образів у реченні, що відповідає синтагмі за допомогою множини типів асоціативних зв'язків:

$$Link - Type = \{r1, r2, r3, r4, r5, r6, r7\}, \tag{15}$$

де *r1* – визначення, *r2* – присудок, *r3* – підмет, *r4* – обставина місця, *r5* – обставина часу, *r6* – обставина, *r7* – додаток, – подальша вербальна деталізація основних типів асоціативних зв'язків може бути здійснена у вигляді підмножин відповідних питальних займенників *Pronoun*.

Словами з множини *Re-Word* будемо позначати атрибути відношень, що мають бути внесеними в блок пам'яті системи.

Як другу основу *Number* будемо розглядати числа таких видів:

$$Number = \{Bi, Integer, Time\}, \tag{16}$$

де *Bi* – n-вимірний бінарний код вузлів булеану;
Integer – цілі невід'ємні числа:

$$Integer = \{x \mid x \in Z^{+}\}; \tag{17}$$

Time – дійсні невід'ємні числа:

$$Time = \{x \mid x \in R^{+}\}. \tag{18}$$

Слід відзначити, що самостійну цінність в запропонованій реляційній моделі мають відношення *Image і Assoc-Twice*, оскільки вони спільно утворюють асоціативну мережу образів і несуть головне змістовне навантаження. Решта всіх відношень фактично моделює стрічку подій або довготермінову пам'ять [6].

Методичне та програмне забезпечення підходу

З метою реалізації окресленого підходу до побудови тезауруса пропонується наступна методика внесення інформації в систему:

1. Користувачу надається можливість внести текст синтагми в систему.
2. Кожному слову синтагми ставиться у відповідність один з образів:
   а) автоматично – у випадку, коли саме таке слово вже існує в пам'яті системи;

b) напівавтоматично – шляхом вибору користувачем зі списку існуючих образів;
c) шляхом ручного введення користувачем нового образу – у випадку відсутності потрібного образу в системі.

3. З метою прив'язки вибраного слова пари до образу у випадку 2b для кожного слова речення складається ранжований список найбільш схожих зовні вербальних позначень з образів *Image*, що вже існують в словнику. При цьому користувач може вибрати в меню існуючу опцію зі списку, при необхідності відкоректувати відповідну статтю словника образів або ввести абсолютно нову.

4. Для введення асоціативних пар синтагми в систему користувачеві надається можливість задати кількість пар, а потім вибрати кожне слово пари в меню, складеного із слів поточного речення Event.

5. Питальний займенник між образами пари задається двома способами:
a) спочатку можна вибрати тип зв'язку з 7 кортежів відношення Link (визначення, присудок, підмет, обставина місця, обставина часу, обставина, доповнення), тоді вибраний тип стає фільтром, і кількість можливих займенників Inter-Pronoun зменшується;
b) якщо спочатку вибрано питання як кортеж з Inter-Pronoun, то для контролю користувачеві автоматично демонструється відповідний йому тип зв'язку з Link, оскільки однакові займенники можуть відноситися до різних типів, наприклад, питання «що?» ставиться як до підмета, так і до доповнення.

З метою поліпшення одночасного доступу багатьох користувачів до словника образів пропонується забезпечити зберігання в одній базі даних не менше двох таблиць *Image* і *Assoc-Twice* як ядра системи, окреме адміністрування бази і організацію до неї паралельного доступу on-Line.

Діючий прототип тезауруса програмно реалізовано на основі технології Python+SQLite, яка поєднує можливості мови запитів SQL до реляційної бази даних з парадигмами об'єктно-орієнтованого і функціонального програмування. Наступний запит до бази даних тестового прикладу в SQLite знаходить найбільш близькі асоціативні зв'язки до одного з образів (у прикладі код образу дорівнює 27):

```
select total(force) as f * from
(select force, b1 as b from a_t where b2=27
union all
select force, b2 as b from a_t where b1=27), image
where bi_i=b group by b order by f desc .
```

У запиті використовується таке представлення a_t:

```
CREATE VIEW a_t as select count(*) as force, b1, b2 from
(select construct_id, bi_i as b1, word_id as w1 from construct as c, words as w where
w.word_id = word_id_1) as a,
(select construct_id, bi_i as b2, word_id as w2 from construct as c, words as w where
w.word_id = word_id_2) as b
where a.construct_id=b.construct_id group by b1,b2 .
```

Аналіз отриманих результатів запиту по контрольному прикладу за текстом О.С. Пушкіна «Казка про золоту рибку» показує, що асоціативні зв'язки між образами-методами «прийти», «кидать», «ловить» та образом-об'єктом «невод» за своєю природою є синтагматичними, а тому пошук за кожною з цих трьох асоціативних пар відразу покаже всі речення, де така пара присутня. Нова якість пошуку проявляється у тому, що його результати залежать виключно від змістовних характеристик синтагм, причому на пошук не впливають синтаксичні особливості речення та морфологічні форми застосованих в ньому слів.

Висновки

У роботі намічені основи одного з можливих підходів до створення «розумного» контента для електронних підручників. На основі вимог концептуальної моделі образного мислення розглянуті принципи побудови глосарія образів, що доповнює традиційний гіпертекст асоціативними зв'язками синтагматичного типу. Пропонується формальна реляційна модель глосарія, що наповнюється в результаті спеціальної обробки текстів з дидактичним матеріалом. Виходячи з обґрунтованого підходу розроблено методику побудови відповідних програмно-педагогічних засобів. Для програмної реалізації навчальної системи обрано технологію Python + SQLite.

Слід зазначити, що нові можливості запропонованого підходу, в тому числі нова якість пошуку, досягаються за рахунок більшої трудомісткості внесення текстів навчального змісту в систему внаслідок розширеного синтаксичного аналізу речень. Проте порівняно невеликі обсяги навчального контенту в системах такого класу не дозволяють вважати дане обмеження критичним.

Література

1. Белянин В.П. Психолингвистика: Учебник. – 5-е изд. – М.: Флинта.
2. Залевская А.А. Значение слова и возможности его описания // Языковое сознание: формирование и функционирование. Сборник статей / Отв. ред. Н.В. Уфимцева. – М., 1998. – С. 35-54.
3. Караулов Ю.Н. Активная грамматика и ассоциативно-вербальная сеть. – М.: ИРЯ РАН, 1999. – 180 с.
4. Бисикало О.В. Подход к моделированию образного механизма оперативной памяти // Труды Второй международной конференции «Новые информационные технологии в образовании для всех: состояние и перспективы развития» (21 – 23 ноября 2007 г.) – Киев, 2007. – С. 336-344.
5. Горошко Е.И. Интегративная модель свободного ассоциативного эксперимента. – М. Харьков: ИЯ РАН; Каравелла, 2001. – 318 с.
6. Бісікало О.В. Алгебраїчна модель лінгвістичного процесора // Информационные технологии в управлении сложными системами: Сборник докладов и тезисов Международной научно-практической конференции (Днепропетровск, 22-23 мая 2008). – Днепропетровск: ИТМ НАНУ и НКАУ, 2008. – С. 23-24.

О.В. Бисикало
Методика построения тезауруса обучающей системы на основе модели образного мышления человека
Работа посвящена созданию «умного» контента для электронных учебников. На основе требований модели образного мышления рассматриваются принципы построения тезауруса образов. Предлагается формальная реляционная модель тезауруса, наполняемого в результате обработки текстов с учебным материалом, а также методика построения программно-педагогических систем такого класса. Действующий прототип системы программно реализован на основе технологии DrScheme + SQLite.

O.V. Bisikalo
Construction Method of Thesaurus of Learning System is on the Basis of Model of Human Figurative Thinking
Work is devoted creation of «clever» content for electronic textbooks. On the basis of requirements of model of image thought principles of construction of thesaurus of appearances are examined. The formal relation model of thesaurus, filled as a result of manipulation of texts with teaching material is offered. Construction method of class of learning systems is worked. The operating prototype of the system is programmatic realized on the basis of technology of DrScheme + SQLite.

Стаття надійшла до редакції 19.08.2008.